

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1. TINJAUAN PUSTAKA

Konsep penambangan data saat ini banyak digunakan pada banyak bidang misalnya Karaolis, et.al (2010) menggunakan konsep penambangan data untuk mengukur faktor resiko awal pada penyakit jantung koroner. Gorakavi, P.K (2010) menggunakan penambangan data dengan metode algoritma C4.5 dalam manajemen proyek kualitatif untuk mencapai hasil yang efektif, yang dapat digunakan untuk membangun sebuah pusat pengetahuan pada proyek *Agile* berdasarkan sejarah statistik. Jelinskis dan Lauks (2008) menggunakan penambangan data untuk mengelola kualitas intrinsik pelayanan dalam MPLS (*Multiprotocol Label Switching*).

Penambangan data juga diterapkan dalam bidang farmasi misalnya Ranjan (2007) menggunakan teknik penambangan data dalam mengelola persediaan obat-obatan dan untuk mengembangkan layanan baru pada produk obat-obatan. Kangwanariyaku, et.al (2010)., Suneetha, N, et.al (2010) menggunakan penambangan data pada *magnetocardiograms* untuk memprediksi penyakit jantung . Shah, et.al (2006) menggunakan teknik penambangan data untuk menganalisis efektivitas pengobatan imunoterapi dengan memahami parameter utama dalam memprediksi cara pengobatan pasien.

Pada bidang pendidikan penambangan data digunakan untuk mengklasifikasikan data pelajar menggunakan metode pohon keputusan (Al-Radaideh, et.al, 2006) sedangkan, Delavari dan Somnuk (2008)., Baeplar dan Murdoch (2010)., Usharani dan Chandrasekaran (2010) menggunakan penambangan data pada lembaga perguruan tinggi guna meningkatkan kualitas manajemen dan teknologi sehingga lebih efisien dalam mengelola dan menentukan strategi-strategi pengelolaan manajemen yang baru.

Penambangan data memiliki metode-metode dalam menyelesaikan suatu persoalan. Salah satu metode yang ada dalam konsep penambangan data adalah metode klasifikasi. Klasifikasi-pun memiliki banyak algoritma dan salah satu algoritamanya adalah algoritma C4.5 yang dapat membentuk sebuah pohon keputusan misalnya, Sunjana (2010) menggunakan algoritma C4.5 untuk mengetahui data nasabah yang dikelompokkan dalam kelas lancar dan kelas tidak lancar, sehingga perusahaan dapat mengambil keputusan menerima atau menolak calon nasabah.

Chih-Chiang dan Jiing-Yun (2011) menggunakan klasifikasi C4.5 untuk menangani masalah teknik sumber daya air. Karegowda, et.al. (2010) menggunakan algoritma C4.5 untuk studi perbandingan menggunakan atribut pemilihan rasio laba dan fitur seleksi berdasarkan korelasi sedangkan, Mengle dan Goharian (2009) menggunakan algoritma C4.5 untuk mengukur fitur yang ambigu pada sebuah algoritma pada dokumen digital.

Amorim, et.al (2009) menggunakan metode klasifikasi untuk perbandingan reduksi algoritma yang diterapkan melalui klasifikasi cacat kulit basah-biru. Sedangkan Karpagavalli., et. al, (2009) menggunakan algoritma C4.5 untuk menangani masalah resiko pemberian bius sebelum melakukan operasi (pra operasi) pada pasien.

Didalam klasifikasi algoritma C4.5 terdapat suatu pohon keputusan dan pohon keputusanpun banyak digunakan dalam berbagai bidang seperti, Lim, Se Hun (2006) menggunakan metode pohon keputusan untuk kolaborasi desain pengendalian manajemen rantai pasok. Erray dan Hacid (2006) menggunakan pohon keputusan sebagai usulan dalam mendeteksi kekeliruan pemakaian biaya pada klasifikasi *mammograms* sebagai teknik pembelajaran yang ter-awasi.

Sunjana (2010) menggunakan klasifikasi pohon keputusan untuk aplikasi penambangan data mahasiswa. Wang dan zhang (2009) menggunakan pohon keputusan sebagai metode yang kuantitatif untuk mendiagnosis pada pengobatan tradisional. Singh, et. al, (2007) menggunakan induksi pohon keputusan berdasarkan algoritma C4.5 untuk mendeteksi pemberian obat yang tepat pada fungsi protein manusia. Rocha dan Junior (2010) menggunakan pohon keputusan untuk mengidentifikasi penipuan atau kecurangan dalam bank. Sedangkan Anyanwu dan Sajjan (2009) menggunakan konsep pohon keputusan untuk mengevaluasi kinerja algoritma secara eksperimental yang umum digunakan pada pohon

keputusan serial dalam pengambilan keputusan berdasarkan catatan data sampel (*Statlog set data*).

Dalam pohon keputusan terdapat pula sebuah sistem pendukung keputusan, dimana sistem ini sebagai salah satu metode dalam pengambilan keputusan, dan sistem inipun banyak digunakan dalam berbagai bidang misalnya, Blaszczyk dan Nowak (2009) menggunakan sistem pendukung keputusan untuk menganalisis proyek bangunan sebagai multikriteria pendukung keputusan sedangkan, Vassilev (2008) menggunakan sistem pendukung keputusan untuk masukan data dan informasi tentang preferensi sebagai pengambil keputusan.

Jigui Sun, et.al (2006) menggunakan sistem pendukung keputusan untuk pemilihan rute jalan pada jasa angkut dan pengiriman bahan-bahan mentah. Khoiruddin (2008) menggunakan Sistem pendukung keputusan untuk penentuan kelayakan calon rintisan sekolah bertaraf internasional. Demir dan Bostanci (2010) dan Daube, et.al (2007) menggunakan sistem pendukung keputusan untuk menganalisis manajemen resiko dan dampak potensial dalam biaya pemasukan dan pengeluaran pada lembaga instansi pemerintahan serta membandingkan pembiayaan sebuah proyek dengan memperkirakan keuntungan dan kerugian sebuah proyek. Sedangkan Suprpto dan Wulandari (2006) menggunakan sistem pendukung keputusan dalam melakukan prakualifikasi kontraktor.

Berdasarkan uraian pustaka diatas maka, penulis mengambil topik dengan menggabungkan tiga (3) konsep untuk dijadikan sebagai penelitian

dalam mengembangkan sebuah aplikasi bantu pada sistem pendukung keputusan pemenang tender proyek. Penggunaan gabungan konsep pemilihan kontraktor, penambangan data dan sistem pendukung keputusan dimaksudkan untuk memperoleh hasil akhir yang lebih efektif dan efisien dalam proses pendukung keputusan penentuan pemenang tender proyek dengan menggunakan algoritma C4.5 yang didasari pada data pemenang tender proyek dan data pendukung (Pustaka Yustisia, 2010) yang sesuai.

2.2. LANDASAN TEORI

2.2.1. Evaluasi dan pemilihan kontraktor Tender Proyek

Pada dasarnya perencanaan strategi suatu perusahaan dalam menjalankan bisnis merujuk pada beberapa prosedur untuk mengembangkan dan mencapai sasaran bisnis tertentu yang telah ditentukan. Seperti kontraktor, perencanaan strategi perusahaan dalam persaingan industri jasa konstruksi mengutamakan penekanan pada identifikasi “seperti apakah” suatu proyek bisa dilakukan penawaran dan “bagaimana” menawarkan harga untuk proyek tersebut.

Tahap awal dalam siklus proyek konstruksi adalah tahap perencanaan dan perancangan (Diputra, 2009). Salah satu tahapan yang paling penting dalam konstruksi adalah proses penawaran (Turskis, 2008) dengan mengevaluasi kinerja kerja setiap kontraktor guna mendapatkan informasi akan kemampuan suatu perusahaan kontraktor dalam mengikuti tender proyek.

Evaluasi terhadap perusahaan kontraktor merupakan bagian yang penting dalam manajemen proyek berdasarkan resiko manajemen (Turkis, 2008., Suputra dan Wiranatha, 2009) yang meliputi perancangan, konstruksi, pembiayaan, operasi, dan pemeliharaan (Daube.et.al, 2007) dalam mengikuti tahapan prakualifikasi.

Prakualifikasi adalah proses penilaian dan pengevaluasian kontraktor oleh pihak pengadaan tender proyek yang di dalamnya termasuk menentukan kriteria-kriteria yang dibutuhkan untuk keberhasilan dalam penyelesaian proyek (Suprpto dan Wulandari, 2006). Prakualifikasi bertujuan untuk mengetahui kemampuan secara teknik dari kontraktor melalui tahap penawaran yang dilakukan oleh perusahaan pengada tender proyek dalam memenuhi syarat-syarat evaluasi sehingga dapat diberikan kepercayaan oleh perusahaan bahwa kontraktor yang nantinya dipilih dapat menyelesaikan proyek dengan baik (Turskis 2008., Suprpto dan Wulandari, 2006). Dengan demikian, tahapan prakualifikasi dapat didefinisikan sebagai Sistem Pendukung Keputusan untuk memberikan keputusan dalam penanganan resiko biaya dan waktu (Błaszczuk dan Nowak, 2009).

2.2.2. Sistem Pendukung Keputusan

Sistem pendukung keputusan (SPK) adalah bagian dari sistem informasi berbasis komputer (termasuk sistem berbasis pengetahuan (manajemen pengetahuan)) yang dipakai untuk mendukung pengambilan keputusan dalam suatu organisasi atau perusahaan. Sistem pendukung keputusan dapat juga dikatakan sebagai sistem komputer yang mengolah data menjadi informasi dalam mengambil keputusan dari masalah semi-terstruktur yang spesifik.

Sistem pendukung keputusan pada tahapan prakualifikasi, pembuatan model *base*-nya dilakukan dengan mengidentifikasi variabel-variabel penentu (Suprpto dan Wulandari, 2006). Tujuan mengidentifikasi variabel-variabel penentu adalah untuk mencapai suatu kriteria penawaran kontraktor dalam menangani sebuah proyek, karena pemilihan kontraktor yang kurang tepat dapat mempengaruhi kualitas (Turskis, 2008., Vassilev, et.al, 2008) dalam menyelesaikan sebuah tender proyek.

Tahapan sistem pendukung keputusan pada pemilihan kontraktor adalah:

1. Definisi masalah.
2. Pengumpulan data atau elemen informasi yang relevan.
3. pengolahan data menjadi informasi baik dalam bentuk laporan, grafik maupun tulisan.
4. menentukan alternatif-alternatif solusi (bisa dalam persentase).

Tujuan dari sistem pendukung keputusan adalah:

1. Membantu menyelesaikan masalah semi-terstruktur.

2. Mendukung manajer dalam mengambil keputusan.
3. Meningkatkan efektivitas bukan efisiensi pengambilan keputusan.

2.2.3. Penambangan data

Penambangan data adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu (Sunjana, 2010) dan juga merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan analisis semi-otomasi dan otomatisasi untuk menemukan data yang sama dan aturan-aturan dari sejumlah besar data dalam database (Kusrini and Luthfi, 2009., Usharani and Chandrasekaran, 2010).

Dalam menemukan sejumlah data yang besar maka, Al-Hegami (2007) menggunakan teknik penambangan data untuk mengklasifikasikan data klasik yang membedakan antara masing-masing kelas sebagai pemecahan kriteria sedangkan, Panda dan Patra (2009) menyatakan pilihan dari setiap algoritma penambangan data adalah suatu kompromi antara waktu yang dibutuhkan untuk membangun model, tingkat deteksi yang rendah dan tingkat alarm palsu serta Indeks Syarat-intrusi.

Konsep penambangan data digunakan untuk membangun sebuah pusat pengetahuan untuk proyek-proyek berdasarkan perencanaan proyek, estimasi biaya, evaluasi kontraktor, dan analisis risiko (Gorakavi, 2010). Dalam konsep penambangan data salah satu metode yang dikenal yakni metode klasifikasi.

2.2.4. Klasifikasi

Klasifikasi merupakan salah satu metode dalam penambangan data yang digunakan untuk menganalisis serangkaian data yang penting untuk setiap kelas yang terdapat dalam kelas yang sama (Al-Hegami, 2007)

Dalam klasifikasi, klasifikasi ini dibangun dari satu set contoh pelatihan dengan label kelas. Metode klasifikasi biasanya menggunakan pengamatan terhadap data yang tersedia sebelumnya untuk mengidentifikasi model yang dapat memprediksi target pengamatan dimasa mendatang.

Definisi melalui pendekatan yang lain diartikan sebagai suatu proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui. Didalam metode klasifikasi ada beberapa algoritma yang sering digunakan dalam memprediksi kelas pada suatu objek, dan salah satu algoritma dari klasifikasi yang digunakan adalah algoritma C4.5.

2.2.5. Algoritma C4.5

Algoritma C4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Hanya saja dalam algoritma C4.5 (Chih-Chiang dan Minegishi, et.al., 2009., Karegowda,et.al, 2010., Al-

Hegami, 2007) pemilihan atribut dilakukan dengan menggunakan *Gain*

Ratio dengan rumus :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots(1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots(2)$$

Di mana :

S = Himpunan kasus

A = Atribut

I = Jumlah Partisi Atribut

|S_i|= Jumlah Kasus pada partisi ke i

|S|= Jumlah Kasus dalam S

Secara umum algoritma C4.5 dapat dibangun menggunakan pohon keputusan (Kusrini and Lutfhi, 2010) dengan cara :

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama

Sebagai contoh, untuk membentuk pohon keputusan dapat dilakukan dengan mencari nilai *Entropy* dan *Gain* terlebih dahulu. Misalkan ditampilkan 14 *dataset* pengujian untuk klasifikasi C4.5 (Kusrini dan Luthfi, 2010) sebagai berikut :

Tabel 1. *Dataset* pengujian klasifikasi C4.5

NO	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Sumber : (Kusrini dan Luthfi, 2010)

Untuk dapat menentukan nilai-nilai *Gain* dan *Entropy* dari masing-masing atribut di atas, maka terlebih dikonversikan data ke dalam bentuk tabel klasifikasi yang lebih detail sebagai berikut :

Tabel 2. Tabel klasifikasi untuk perhitungan nilai *Gain* dan *Entropy*

Node	Kriteria	Kategori	Jumlah Kasus (s)	No (s1)	Yes (s2)	<i>Entropy</i>	<i>Gain</i>
1	Total		14	4	10	0.863	
	Outlook						0.258
		Cloudy	4	0	4	0	
		Rainy	5	1	4	0.721	
		Sunny	5	3	2	0.970	
	Temperature						0.183
		Cool	4	0	4	0	
		Hot	4	2	2	1	
		Mild	6	2	4	0.918	
	Humidity						0.370
		High	7	4	3	0.985	
		Normal	7	0	7	0	
	Windy						0.005
		False	8	2	6	0.811	
		True	6	4	2	0.918	

Sumber : (Kusrini dan Luthfi, E.T. 2010)

Setelah tabel 2 dibuat, maka langkah selanjutnya adalah melakukan perhitungan dengan mencari nilai *Entropy* menggunakan persamaan (1). Setelah menentukan nilai *Entropy* langkah selanjutnya yaitu mencari nilai *Gain* menggunakan persamaan (2).

Langkah pertama yang harus dilakukan yaitu dengan mencari nilai

Entropy total sebagai berikut :

Baris Total Kolom *Entropy* pada Tabel 2 di hitung dengan Rumus :

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

$$Entropy (Total)$$

$$= ((-4/14) * -1,807) + ((-10/14) * (-0,385))$$

$$= 0,516 + 0,346$$

$$= 0,863$$

Langkah selanjutnya, mencari nilai *Entropy* untuk masing-masing atribut, misalnya pada Outlook (Cloudy, Rainy dan Sunny) sebagai berikut:

$$EntropyOutlook(Cloudy) = (-0/4 * \text{Log}_2(0/4) + (-4/4) * \text{Log}_2 (4/4)$$

$$= 0 + 0$$

$$= 0$$

$$EntropyOutlook(Rainy) = (-1/5) * \text{Log}_2 (1/5) + (-4/5) * \text{Log}_2 (4/5)$$

$$= 0,464 + 0,257$$

$$= 0,721$$

$$EntropyOutlook(Sunny) = (-3/5) * \text{Log}_2 (3/5) + (-2/5) * \text{Log}_2 (2/5)$$

$$= 0,442 + 0,528$$

$$= 0,970$$

Perhitungan selanjutnya adalah menghitung nilai *Gain* total untuk Outlook dengan menggunakan persamaan (2) berdasarkan nilai *Entropy* dari masing-masing atributnya sebagai berikut :

$$Gain (S, A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i)$$

Information *Gain* (Total, Outlook)

$$\begin{aligned}
 &= 0,863 - ((4/14 * 0) + (5/14 * 0,721) + (5/14 * 0,970)) \\
 &= 0,863 - (0 + 0,240 + 0,3467) = 0,863 - 0,604 \\
 &= 0,258
 \end{aligned}$$

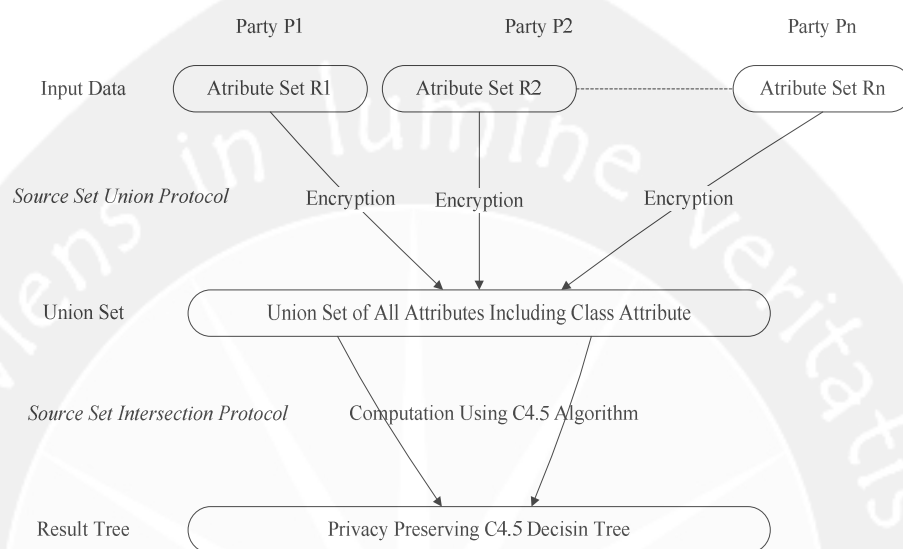
Perhitungan nilai *Entropy* dan *Gain* dilakukan sampai atribut terakhir sehingga dapat diperoleh nilai-nilai *Gain* dan *Entropy* tertentu. Hasil dari nilai *Gain* kemudian dibandingkan dengan nilai *Gain* lainnya untuk mencari nilai tertinggi sehingga *Gain* yang merupakan nilai tertinggi akan dijadikan acuan untuk melakukan proses perhitungan kembali pada proses pembentukan struktur pohon keputusan (Kusrini dan Luthf, 2010).

2.2.6. Pohon Keputusan

Salah satu metode Penambahan data yang umum digunakan adalah Pohon Keputusan. Pohon Keputusan adalah struktur *flowchart* yang menyerupai *tree* (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada Pohon Keputusan di telusuri dari simpul akar ke simpul daun sehingga prediksi kelas pada pohon keputusan mudah untuk dikonversi ke aturan klasifikasi (*classification rules*).

Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*break* down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan

lebih menginterpretasikan solusi dari permasalahan. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi. Sebagai contoh dapat di lihat pada gambar 2.2.6.1.



Gambar 2.2.6.1. *Proposed privacy preserving decision tree*

(Gangrade dan Patel, 2009)

Pembentukan pohon keputusan seperti gambar 2.2.6.1. diatas pada dasarnya dalam pembentukan sebuah pohon, langkah awal yang harus ditentukan adalah menentukan Atribut Set dari masing-masing kelas atribut. Setelah semua atribut ditentukan, atribut-atribut yang ada dalam masing-masing kelas tersebut diuji menggunakan algoritma C4.5 sehingga hasil yang diharapkan dapat membentuk sebuah pohon keputusan dan dapat memberikan suatu informasi atau pengetahuan baru sebagai dasar pengambilan keputusan.